(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
14 March 2002 (14.03.2002)

PCT

(10) International Publication Number
WO 02/21331 A1

(51) International Patent Classification[7]: G06F 17/30

(21) International Application Number: PCT/GB01/03787

(22) International Filing Date: 22 August 2001 (22.08.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
0021881.8 6 September 2000 (06.09.2000) GB

(71) Applicant *(for all designated States except US)*: **ARGO INTERACTIVE GROUP PLC** [GB/GB]; Oak House, Shackleford Road, Elstead, Surrey GU8 6LB (GB).

(72) Inventor; and

(75) Inventor/Applicant *(for US only)*: **SPOONER, Roger, Ian, William** [GB/GB]; 30 South Oswald Road, Edinburgh, Scotland EH9 2HG (GB).

(74) Agents: **ROBINSON, Nigel, Alexander, Julian** et al.; D. Young & Co., 21 New Fetter Lane, London EC4A 1DA (GB).

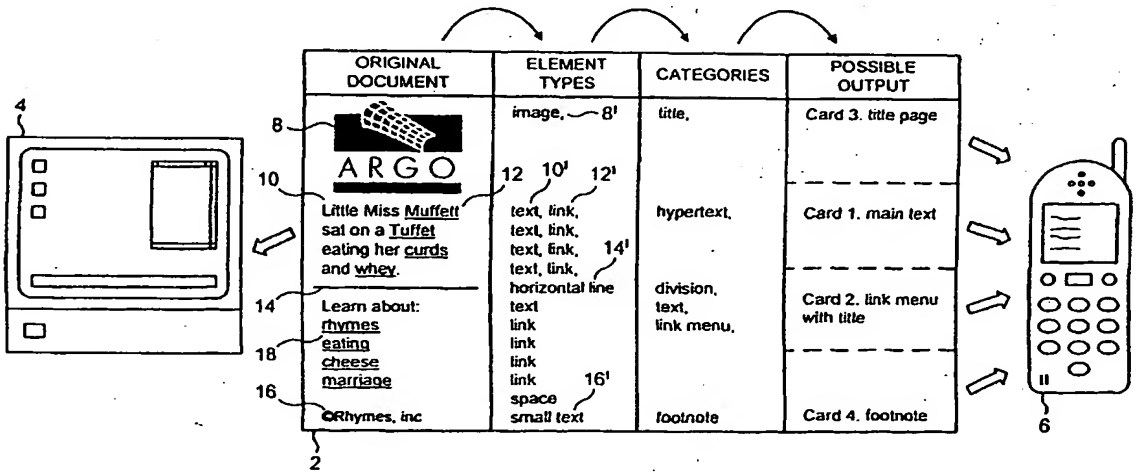
(81) Designated States *(national)*: JP, US.

(84) Designated States *(regional)*: European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR).

Published:
— *with international search report*
— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: ANALYSING HYPERTEXT DOCUMENTS



(57) Abstract: A hypertext documet (2) is analysed by identifying document elements within it and then categorising those document elements to given element types. Heuristic pattern matching is then performed upon the categorised element types to identify patterns indicative of different document regions. The original document may then be divided into separate documents based upon the identified document portions.

WO 02/21331 A1

BNSDOCID: <WO___0221331A1_I_>

ANALYSING HYPERTEXT DOCUMENTS

This invention relates to the field of the analysis of hypertext documents. More particularly, but not exclusively, this invention relates to the type of analysis of hypertext documents that is useful in transcoding such documents from a form suitable for display on a conventional personal computer to a form suitable for display on a mobile telephone or other display device having a more limited display capability.

Internet web page transcoders attempt to reduce page content, summarise, divide, or otherwise process a web page on the basis of its content with the aim of making the material more suitable for display upon devices for which that material was not originally intended. Currently, transcoding systems operate using very localised HTML element processing. However, there is a limit to the degree of sophistication in the processing that can be performed at an individual element level and this impedes the improvement in intelligence and functionality of transcoding systems.

Some transcoders operate as a "proxy" on a computer on a network which can be contacted by the end user's device in place of it contacting the computer holding the desired document. Other transcoding systems can operate as "plug-ins" to computer software which dispatches documents from the original computer. It is also possible that the transcoding system could reside on the user's computer that receives the document.

Viewed from one aspect the present invention provides a method of identifying one or more regions of a hypertext document formed of a plurality of document elements, said method comprising the steps of:

parsing said plurality of document elements to identify an element type for each of said plurality of document elements;

categorising one or more patterns of element types within said hypertext document indicative of respective categorised regions of said hypertext document; and

selecting one or more categorised regions to be identified as said one or more regions of said hypertext document in dependence upon a confidence measure associated with each categorised region.

The invention recognises that a more abstract view of a web page being processed can be used to identify the structure of that web page in a manner that assists

BNSDOCID: <WO 0221331A1 >

subsequent processing. In particular, the invention steps away from the analysis of individual elements and instead parses the individual elements for element type and looks for patterns amongst those elements types. This allows different regions of a hypertext document to be identified in a way that was not previously possible.

Whilst the identification of different regions of a hypertext document may be useful for various different reasons, it is particularly advantageous when it is desired to divide an import hypertext document into smaller processed documents as the invention allows more intelligent and usable divisions between the processed documents to be achieved.

Whilst strictly deterministic algorithms could be used for the pattern matching, it is preferable to use heuristic algorithms as these are more able to cope with the wide variety of different hypertext document layouts that can be encountered whilst still accurately identifying the different regions of a document.

Whilst the pattern matching used could take a variety of forms, a particularly efficient form of pattern matching is one in which each document element is compared with its neighbours to see if there is a consistent pattern of relationship between them at which point these matching elements may be grouped together and additional new neighbours considered. Accordingly, the group of elements matching a given pattern will grow by sequentially including neighbouring elements until neighbouring elements are encountered that no longer fit the pattern.

Additionally, patterns between neighbouring or non-adjacent regions may also be subsequently matched.

A further characteristic of document element that has been found useful in identifying document regions is the relative position of a document element within a hypertext document as a whole. As an example, images at the top of a document are often associated with a title of a document and small text at the bottom of a document is often associated with a footnote.

The technique of the present invention has been found to be particularly effective in identifying mixed blocks of text and links as hypertext prose, blocks of adjacent neighbouring links as link menus, text neighbouring a link menu as a title for the link menu, an image close to the top of a document as part of a title, small text

BNSDOCID: <WO 0221331A1 I >

toward the bottom of a document as part of a footnote and a horizontal line spaced away from any images or graphics elements as a division within a document.

It will be appreciated that the technique of the present invention may be used in the analysis of many different types of hypertext document, e.g. documents containing both text and additional elements such as images, graphics and links. The invention is particularly well suited to the analysis of HTML, WML and CHTML documents but could be applied to other hypertext document formats such WML documents or even PDF documents and the like.

Viewed from another aspect, the present invention provides apparatus for identifying portions of a hypertext document formed of a plurality of document elements, said apparatus comprising:

parsing logic operable to categorise said plurality of document elements to identify an element type for each of said plurality of document elements;

categorising logic operable to identify one or more patterns of element types within said hypertext document indicative of respective regions of said hypertext document; and

selecting logic operable to select one or more categorised region to be identified as said one or more regions of said hypertext document in dependence upon a confidence measure associated with each categorised region.

The invention also provides a computer programme storage medium for storing a computer programme to control a general purpose computer to operate in accordance with the above described techniques. The computer storage medium could be a CD, a hard disk drive or a downloaded computer file.

An embodiment of the invention will now be described, by way of example only, with reference to the accompanying drawings in which:

Figure 1 schematically illustrates the technique of the present invention in analysing a hypertext document;

Figure 2 is a simplified flow diagram illustrating the processing performed in the technique in Figure 1;

Figures 3A and 3B are a table indicating the correspondence between patterns of element types, element positions, and the portions of a document to which these correspond and a "regular expression" view of the same relationships;

BNSDOCID: <WO 0221331A1 I >

Figure 4 is a flow diagram illustrating in more detail how the pattern matching may be performed;

Figure 5 is a schematic illustration of a data processing apparatus that may be used to implement the techniques of the present invention;

Figures 6 and 7 illustrate a source hypertext document viewed at element type and region levels; and

Figure 8 illustrates various example patterns which may be matched.

Figure 1 schematically illustrates the technique of one example of the present invention. An input hypertext document in the form of a HTML page 2 is the starting point for the process. This HTML page 2 has been designed by its author to be suited for display upon a computer monitor 4 of a typical personal computer. If it is desired to access this HTML page, or at least the content of that page, using a mobile telephone 6, or other small display screen device such as a personal digital assistant, then the original form of the HTML document 2 is inappropriate for display upon the mobile telephone 6. More particularly, the mobile telephone 6 will have a small and less capable display than is provided by the computer monitor 4 and accordingly only a small portion of the contents of the HTML page 2 could be fully displayed at any given time.

For this reason, transcoding products seek to reduce, sub-divide and simplify the content of source pages in a manner that enables them to be better displayed upon less capable (or just different) display devices. The transcoding processing may take place in a proxy server disposed between the user device and the source server or it may take place in the source server itself.

The first step in the analysis of the HTML document is to identify the different element types within that document. Accordingly, the image 8 is identified as an image type 8′, the text 10 is identified as a text type 10′, the link 12 is identified as a link type 12′ and so forth. The horizontal line 14 is identified as a horizontal line type 14′ and the footnote text 16 is identified as a small text type 16′. Thus, the first step in the processing of the HTML document achieves an indication of each document element as a particular element type. This might be represented as a "parse tree" as in Figure 6.

BNSDOCID: <WO___0221331A1_I_>

The next stage in the processing is to perform pattern matching upon the element types to find different regions of the original HTML document 2 (this may use heuristic algorithms). In particular, the image 8 and its associated image type 8′ positioned towards the top of the HTML document 2 is identified as part of a title portion. Any text immediately preceding the image 8 or following the image 8 and spaced from other content within the HTML page 2 may also be grouped together with the image 8 and identified as part of the title.

The mixed section of text 10 and links 12 is pattern matched to a block of hypertext prose. Such pattern matching may be achieved by starting with the initial document element and then comparing it with its neighbours to identify a pattern consistent with a particular type of document portion and growing that portion by encompassing further neighbouring elements until the identified pattern no longer holds true. A block of hypertext prose will typically form the main prose of an HTML page and represent highly significant information content. Accordingly, in a transcoding system such a portion may be identified as the most significant to a user and presented first to that user. Content containing elements rather than layout elements are given a greater weighting in assessing user importance.

The horizontal line 14 within the HTML page 2 is identified as a horizontal line type 14′. As this horizontal line type 14′ is spaced apart from any images or graphical images within the HTML page 2, it is pattern matched to represent a division within the HTML page 2. It has been found that if a horizontal line is close to or neighbouring an image element or a graphical element, then it is more usually intended by the author of the HTML page 2 to form part of that image or graphic (e.g. part of the border) and accordingly if such a disposition is detected then the horizontal line will not be pattern matched to represent a division.

Tags indicative of the document structure per se (e.g. paragraph tags) are given their original meaning in deciding how to divide up a page.

A sequence of adjacent hypertext links 18 is pattern matched to represent a link menu. If such a link menu is closely preceded or followed by a text element, then that text element is taken to represent a title or footnote to the link menu.

Small text identified towards the bottom of a hypertext page 2 or close to an image 8 is identified as being a footnote region or caption . The confidence in the

BNSDOCID: <WO 0221331A1 I >

identification of such small text as a footnote region is determined as higher if the small text is spaced from preceding elements within the HTML page 2.

Once all the different document regions have been identified, then the original content of the HTML page 2 may be divided into processed documents (cards) that may be separately supplied to and displayed by a mobile telephone 6. A division (or structural markup element) identified within the HTML page 2 is taken as a strong indication for a point at which the original content can be divided. The order in which the processed documents will be displayed to a user may be selected in dependence upon the nature of the document portion that has been pattern matched, e.g. a main text of mixed hypertext prose and a link menu will be regarded as more significant and more desirable to display to a user than would be a title page or a footnote page.

The patterns matched may be indicative of more than one way in which the document may be divided into regions. All possible ways in which patterns found may identify regions are recorded with associated confidence measures. When the patterns have all been matched then the regions to be used are selected as those having the highest confidence measures . A given document element may lie in two or more possible regions, but these highest confidence region will be used.

Figure 2 is a simplified flow diagram of the processing that may be performed as part of a transcoding process utilising the present invention. At step 20 a web page is received by the system. At step 22 the web page is parsed to identify different document elements within it. The markup language form of a web page makes it relatively straight forward to identify different document elements by their associated tags. At step 24 each of the identified document elements is noted as a particular-type of element, for example the HTML note would be identified as "small text". At step 26 a search is made through the element types to identify patterns characteristic of different regions of a document. This pattern matching may use heuristic algorithms with associated confidence levels for the pattern matching achieved in accordance with known pattern matching techniques. At step 28, the received web page is divided into a number of separate processed documents in dependence upon the identified document regions with the highest confidence measures as produced by step 26.

BNSDOCID: <WO__0221331A1_I_>

The division on an input web page in accordance with the present technique may be coupled with other transcoding techniques to reduce the content of a web page to focus upon the content useful to a user and adapt the web page to a form more suited for manipulation using a device other than that for which the web page was originally designed. The technique of the present invention is strongly advantageous as part of a transcoding system as a whole.

Figure 3A is a table indicating a relationship between element types, element positions and corresponding regions. The pattern matching discussed above is responsive to the element types and element positions to heuristically identify corresponding document regions. The pattern matching performed may associate a confidence level with each pattern matched. Accordingly, a particular document element might form part of two possible matching patterns indicative of different document portions. The confidence levels associated with the pattern matches can be used to select which of the pattern matches forms the basis of subsequent processing and the division of the document in accordance with the identified regions. Patterns between identified regions may also be pattern matched to identify larger regions.

Figure 3B corresponds to Figure 3A, but in this case gives a regular expression view of the relationships between elements type that may be pattern matched. Where several exits are shown from a stage, any may be taken depending upon the next element type to be processed with the source document.

Figure 4 is a flow diagram schematically illustrating the processing performed in the pattern matching. At step 30, an initial document element within an input HTML page is selected. This will typically be the first document element in the page. Step 32 compares the selected element type, position and any neighbouring element types to determine if a match to one of a plurality of predetermined patterns is indicated with a sufficient degree of confidence. If such a match is indicated, then processing proceeds to step 34 at which an attempt is made to extend the region to include further neighbouring elements that would also match. Accordingly, the region of a document matching a given pattern is grown by successive processing in step 32 and 34 until the matching region can be extended no further. At this stage processing proceeds to step 34.

BNSDOCID: <WO 0221331A1 >

Step 34 checks the confidence level associated with the pattern match achieved in the preceding steps to determine whether this has achieved identification of a document region with a greater or lesser degree of confidence than any preceding match identified for the elements within that document region. Steps 36 and 38 serve to either establish the newly matched region as the current preferred candidate or not in dependence upon the comparison of the confidence levels at step 34.

Step 40 checks to see whether the last document element within the hypertext document has already been used as a starting point for the pattern matching process. If this is not the case, then processing proceeds via step 42 at which the next document element is selected as a starting point than that which previously formed the starting point for the entry into the pattern matching steps 32 and 34. Thus, each document element in term is used as a starting point for pattern matching until all of the document elements have been so used. The highest confidence level patterns and portions resulting when the processing has finished are those used for subsequent transcoding operations, such as page division. Subsequently, identified regions may be used as the starting point for further region/region or region/element matching.

Figure 5 schematically illustrates a data processing system of the type which may perform the technique of the present invention. This data processing system includes a central processing unit 44, a read only memory 46, a random access memory 48, a hard disk drive 50 and a network link 52. Such a general purpose data processing system will execute a computer program that may be stored upon the hard disk drive 50, within the read only memory 46 or downloaded via the network link 52. The working memory during such program execution will be provided by the random access memory 48. The results of such data processing may be displayed to a user of another device with which the system communicates through the network link 52. The user can give commands to the system via the user input/output unit 58 in conjunction with the keyboard 60 and the mouse 62. It will be appreciated that the central processing unit 44 executing computer program instructions effectively serves as logic for performing the processing steps described above. The computer program executed by the data processing system may be loaded into the system via a tangible medium, such as a compact disk or floppy disk, or downloaded via the network link 52.

BNSDOCID: <WO__0221331A1_I_>

Figure 6 illustrates the structure of the document of Figure 1 in the form of the element types identified and the structural markup tags that may be included. Figure 7 is a view of the same page at a higher level of abstraction once the document regions have been identified.

Figure 8 illustrates various document element and region patterns that may be matched to predefined criteria indicative of certain document regions. Many further patterns are possible.

BNSDOCID: <WO 0221331A1 I >

CLAIMS

1. A method of identifying one or more regions of a hypertext document formed of a plurality of document elements, said method comprising the steps of:

parsing said plurality of document elements to identify an element type for each of said plurality of document elements;

categorising one or more patterns of element types within said hypertext document indicative of respective categorised regions of said hypertext document; and

selecting one or more categorised regions to be identified as said one or more regions of said hypertext document in dependence upon a confidence measure associated with each categorised region.

2. A method as claimed in claim 1, further comprising the step of dividing said hypertext document into a plurality of processed documents in dependence upon said regions of said hypertext document identified by said one or more patterns of element types.

3. A method as claimed in claim 2, wherein different regions of said hypertext document appear in respective different processed documents.

4. A method as claimed in any one of the preceding claims, wherein said step of categorising one or more patterns uses heuristic algorithms.

5. A method as claimed in any one of the preceding claims, wherein said step of categorising one or more patterns starts with each document element and compares an element type for that document element with element types for neighbouring document elements to identify a pattern of neighbouring element types indicative of a region of said hypertext document.

6. A method as claimed in claim 5, wherein said step of categorising one or more patterns starts with each of any hypertext links within said hypertext document.

BNSDOCID: <WO__0221331A1_I_>

7. A method as claimed in claim 5, wherein said step of categorising one or more patterns starts with each of any textual headings within said hypertext document.

8. A method as claimed in claim 5, wherein said step of categorising one or more patterns starts with each of any graphical images within said hypertext document.

9. A method as claimed in any one of claims 5 to 8, wherein said comparison is made with previously categorised regions of said hypertext document.

10. A method as claimed in claim 9, wherein said comparison of categorised regions is made between non-consecutive regions of said hypertext document.

11. A method as claimed in any one of the preceding claims, wherein said step of categorising one or more patterns combines one or more of the methods claims 5, 6, 7 and 8.

12. A method as claimed in any one of the preceding claims, wherein said step of categorising one or more patterns is responsive to a relative position of a document element within said hypertext document as a whole when identifying a portion of said hypertext document.

13. A method as claimed in any one of the preceding claims, wherein a plurality of possible categorised regions may correspond to a given document element and a confidence measure is associated with each of said plurality of possible categorised regions.

14. A method as claimed in claim 13, wherein, after all overlapping possible categorised regions have been found, overlapping categorised regions are compared and a categorised region having a highest confidence level is selected from amongst other categorised regions that overlap its document elements.

BNSDOCID: <WO 0221331A1 I >

15. A method as claimed in any one of the preceding claims, wherein a mixed block of text and links is identified as hypertext region.

16. A method as claimed in any one of the preceding claims, wherein a plurality of neighbouring links is identified as a link menu region.

17. A method as claimed in claim 16, wherein text immediately preceding said link menu region is identified as a title for said link menu region.

18. A method as claimed in any one of the preceding claims, wherein an image close to the top of said hypertext document is identified as part of a title portion.

19. A method as claimed in any one of the preceding claims, wherein text close to the bottom of said hypertext document and with a font size smaller than normal for said hypertext document is identified as part of a footnote portion.

20. A method as claimed in any one of the preceding claims, wherein a horizontal line spaced away from any images or graphics within said hypertext document is identified as a division.

21. A method as claimed in any one of the preceding claims, wherein a textual heading is identified as a division.

22. A method as claimed in any one of the preceding claims, wherein a boundary between table cells is identified as a division.

23. A method as claimed in claim 2 and any one of claims 20 to 22, wherein said division is used to identify where said hypertext document should be split into processed documents.

BNSDOCID: <WO___0221331A1_I_>

24. A method as claimed in any one of the preceding claims, wherein a structural markup element is recognised as its original meaning.

25. A method as claimed in any one of the preceding claims wherein a value of a relative importance to a user of each categorised region is allocated for later processing.

26. A method as claimed in any one of the preceding claims, wherein said method of categorising regions is part of a method for converting a hypertext document originally intended for display on a first type of display device into a form adapted for display on a second type of display device.

27. A method as claimed in claim 26, wherein said first type of display device is a conventional computer monitor.

28. A method as claimed in any one of claims 26 and 27, wherein said second type of display device is one of: a mobile telephone display and a personal digital assistant display.

29. A method as claimed in any one of the preceding claims, wherein said hypertext document is an HTML document.

30. A method as claimed in any one of claims 1 to 29, wherein said hypertext document is one of a WML document and a CHTML document.

31. A method as claimed in any one of the preceding claims, further comprising the steps of employing iterative growth and combination of categorised regions of said document so as to reach one of a preferred number of said regions and a preferred size of each said region.

BNSDOCID: <WO___0221331A1_I_>

32. A method as claimed in any one of the preceding claims, further comprising the step of dividing the said hypertext document into a plurality of processed documents in dependence upon a size of said categorised regions.

33. A method as claimed in any one of the preceding claims, further comprising the step of dividing the said hypertext document into a plurality of processed documents in dependence upon a measure of importance to a user of said categorised regions.

34. A method as claimed in any one of the preceding claims, further comprising the step of dividing the said hypertext document into a plurality of processed documents in dependence upon a predetermined maximum size for said plurality of processed documents matched to a target display device size.

35. A method as claimed in any one of the preceding claims, wherein said method steps are performed in real time upon an access request to said hypertext document.

36. A method as claimed in any one of the preceding claims, wherein said steps of parsing, categorising and selecting are performed on a computer acting as a proxy between a user device requesting said hypertext document and a server storing said hypertext document.

37. A method as claimed in any one of claims 1 to 34, wherein said steps of parsing, categorising and selecting are performed on a server storing said hypertext document prior to dispatch of a modified form of said hypertext document to a user device.

38 Apparatus for identifying portions of a hypertext document formed of a plurality of document elements, said apparatus comprising:

parsing logic operable to categorise said plurality of document elements to identify an element type for each of said plurality of document elements;

BNSDOCID: <WO__0221331A1_I_>

categorising logic operable to identify one or more patterns of element types within said hypertext document indicative of respective regions of said hypertext document; and

selecting logic operable to select one or more categorised region to be identified as said one or more regions of said hypertext document in dependence upon a confidence measure associated with each categorised region.

39. A computer program storage medium for storing a computer program to control a computer to perform a method as claimed in any one of claims 1 to 36.

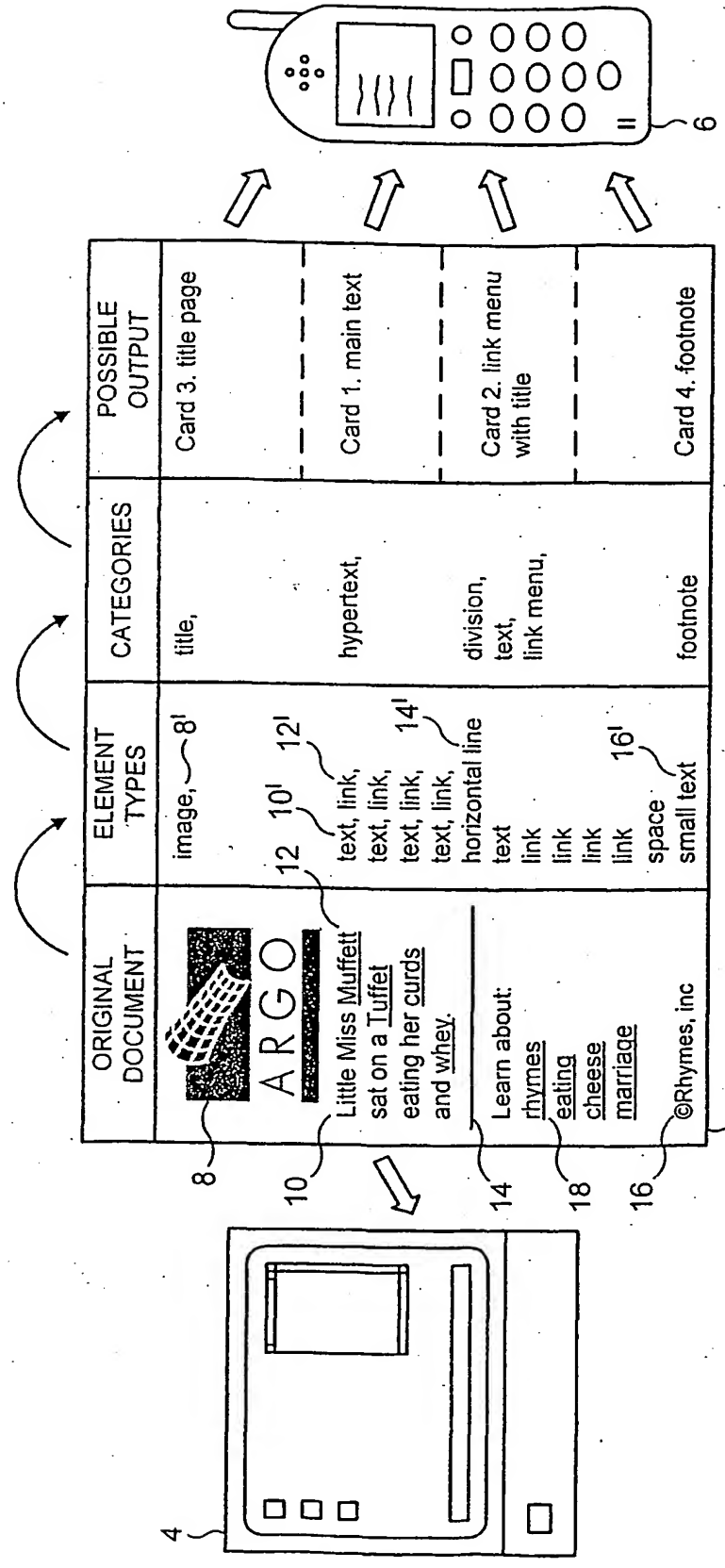


FIG. 1

BNSDOCID: <WO___0221331A1_I_>

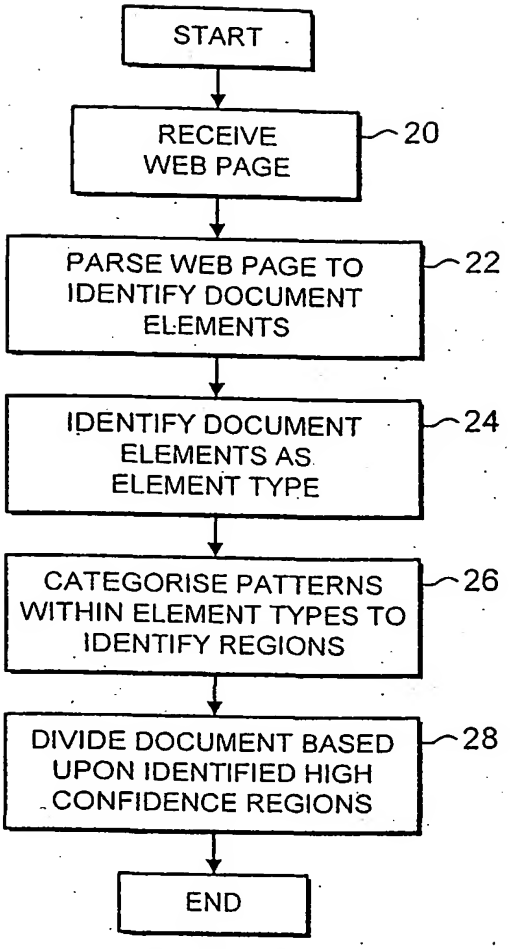


FIG. 2

	ELEMENT TYPE(S)	ELEMENT(S) POSITION	REGION
A.	MIXED TEXT AND LINKS	ANYWHERE	HYPERTEXT BLOCK
B.	NEIGHBOURING LINKS	ANYWHERE	LINK MENU
C.	TEXT	IMMEDIATELY ABOVE A LINK MENU	LINK MENU TITLE
D.	IMAGE	CLOSE TO TOP	TITLE
E.	HORIZONTAL LINE	SPACED FROM ANY IMAGES OR GRAPHICS	DIVISION
F.	SMALL TEXT	CLOSE TO BOTTOM OR TITLE	FOOTNOTE
G.	IMAGE	SURROUNDED BY HORIZONTAL LINES	IMAGE

FIG. 3A

BNSDOCID: <WO___0221331A1_I_>

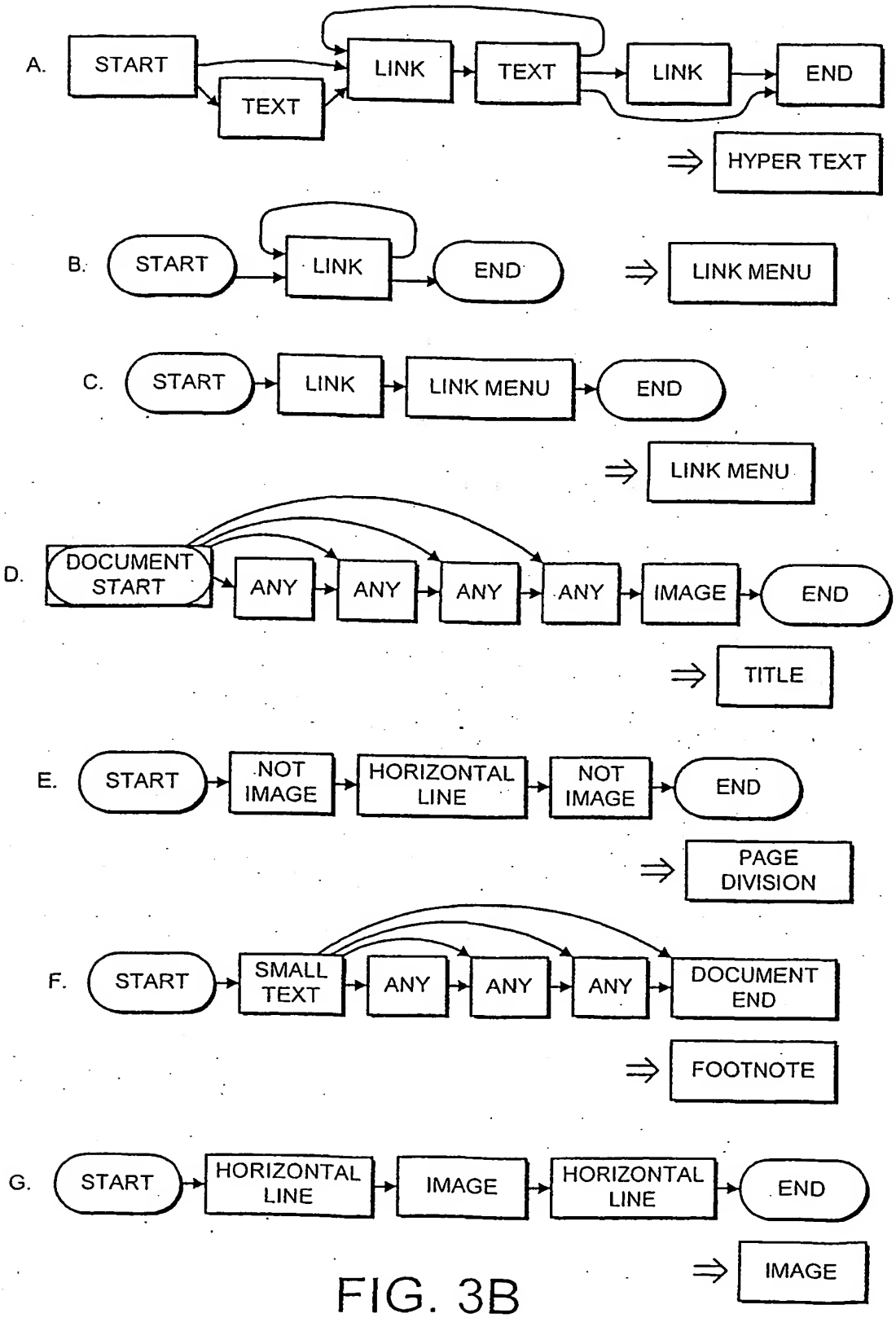


FIG. 3B

BNSDOCID: <WO___0221331A1_I_>

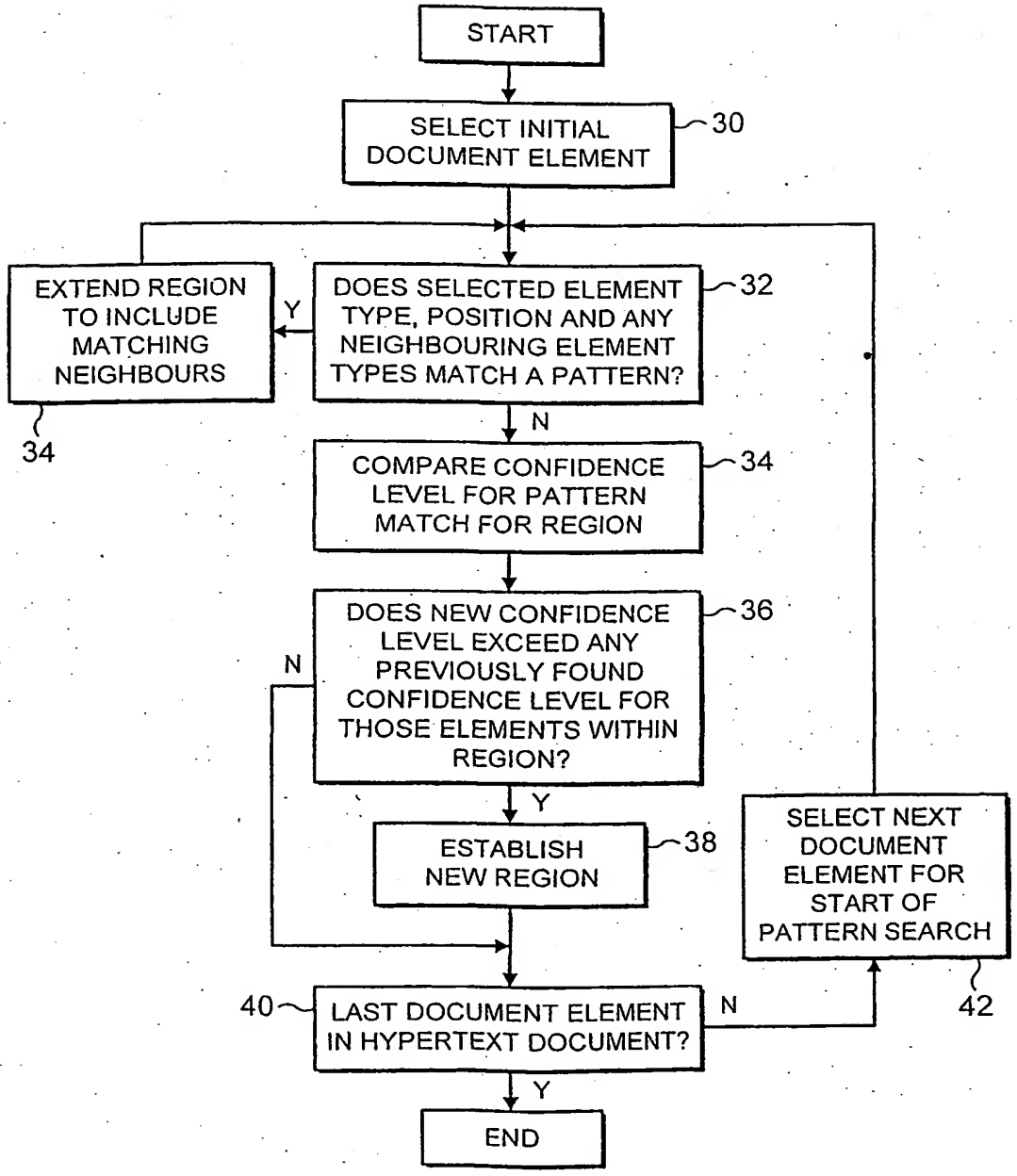


FIG. 4

BNSDOCID: <WO 0221331A1 I >

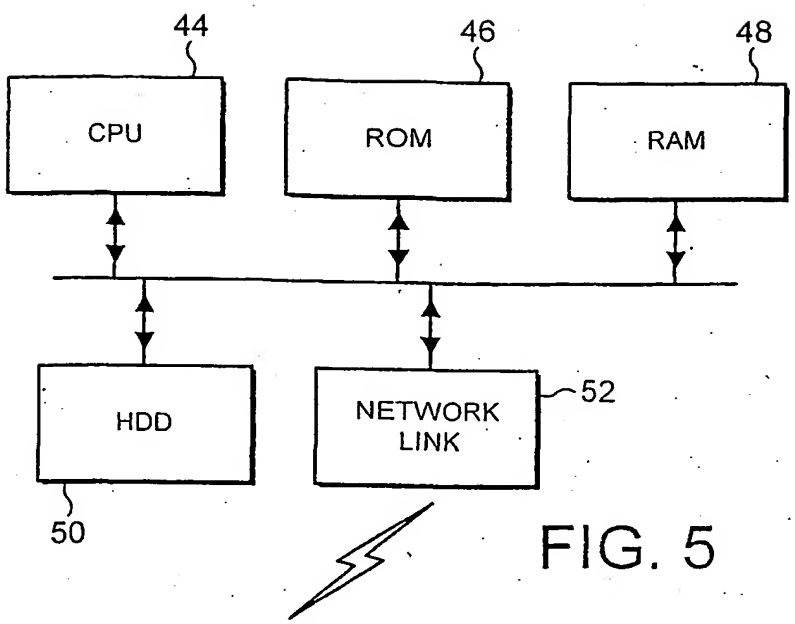


FIG. 5

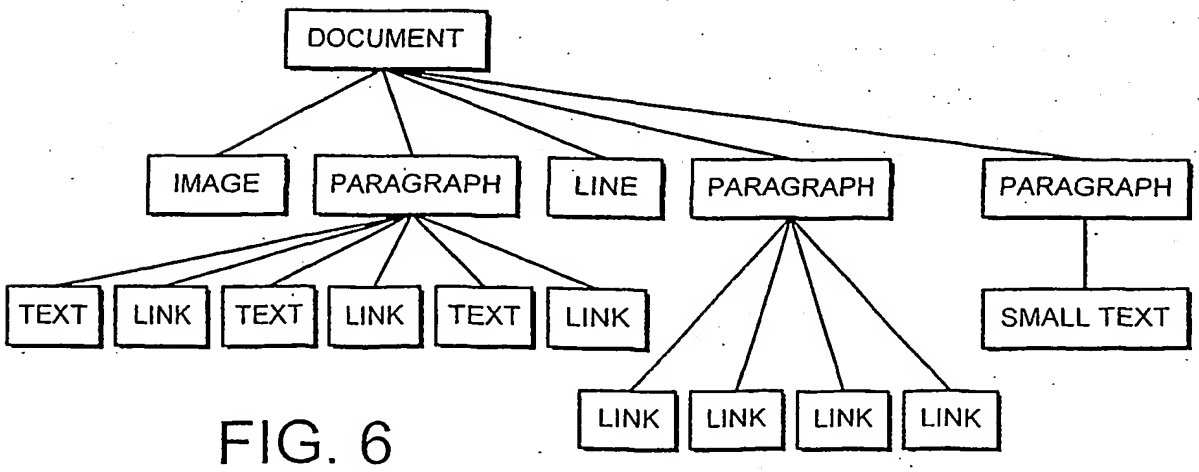


FIG. 6

NSDOCID: <WO___0221331A1_I_>

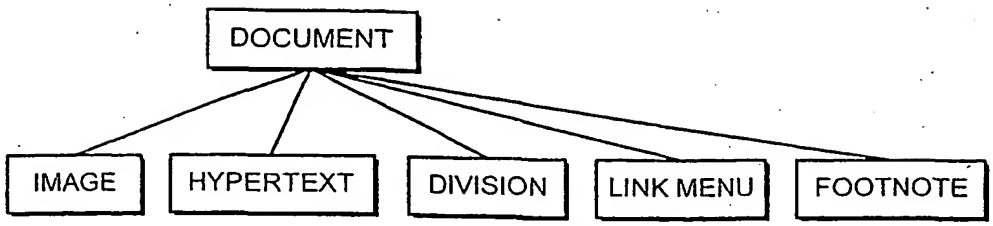


FIG. 7

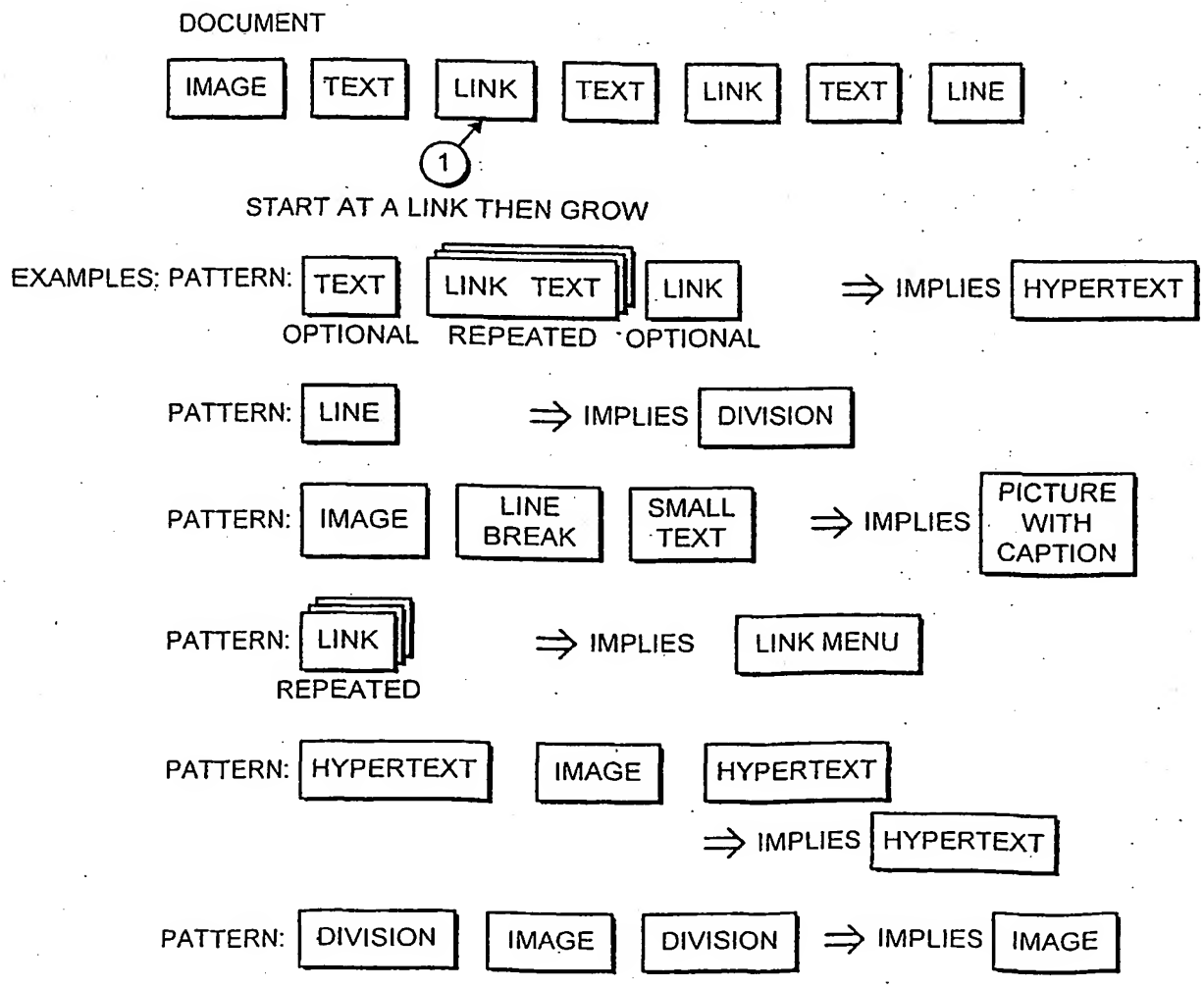


FIG. 8

BNSDOCID: <WO 0221331A1 I >

INTERNATIONAL SEARCH REPORT

Internatl Application No
PCT/GB 01/03787

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 G06F17/30

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7 G06F

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)
EPO-Internal, WPI Data, PAJ, INSPEC, IBM-TDB

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	GB 2 339 374 A (IBM) 19 January 2000 (2000-01-19) abstract; figure 1 page 12, line 24 -page 13, line 33; figures 3,6 page 14, line 23 -page 15, line 32; figures 8,10 page 18, line 34 -page 19, line 43; figures 9,13 page 26, line 18 -page 27, line 6; figure 16 --- -/--	1-4,12, 13, 25-30, 32-39

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

° Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
14 February 2002	21/02/2002

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Polzer, A

Form PCT/ISA/210 (second sheet) (July 1992)

BNSDOCID: <WO___0221331A1_I_>

INTERNATIONAL SEARCH REPORT

Internati Application No
PCT/GB 01/03787

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication,where appropriate, of the relevant passages	Relevant to claim No.
A	WO 00 39666 A (CARLINO KENNETH F; HOHL KENNETH B (US); SPYGLASS INC. (US)) 6 July 2000 (2000-07-06) abstract; figure 1 page 36, line 3 -page 37, line 11; figures 2-4; table 6 page 38, line 12 -page 44, line 18; figures 5-8; tables 7-9 ---	1,38,39
A	BICKMORE T ET AL: "WEB PAGE FILTERING AND RE-AUTHORING FOR MOBILE USERS" COMPUTER JOURNAL, OXFORD UNIVERSITY PRESS, vol. 42, no. 6, 1999, pages 534-546, XP000920338 SURREY, GB ISSN: 0010-4620 section 3.1; figures 1-3 page 538, right-hand column, line 8 - line 28 section 3.3.2; figures 4,6 section 3.3.3 ---	1,38,39
A	KAASINEN E ET AL: "Two approaches to bringing Internet services to WAP devices" COMPUTER NETWORKS, ELSEVIER SCIENCE PUBLISHERS B.V., vol. 33, no. 1-6, June 2000 (2000-06), pages 231-246, XP004304769 AMSTERDAM, NL ISSN: 1389-1286 section 4.1; figures 1,3 section 4.2 ---	1,38,39
A	WO 00 42531 A (YAHOO INC) 20 July 2000 (2000-07-20) page 6, line 7 -page 8, line 8; figures 3-6 -----	1,38,39

BNSDOCID: <WO___0221331A1_I_>

INTERNATIONAL SEARCH REPORT

Internal Application No
PCT/GB 01/03787

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
GB 2339374 A	19-01-2000	US 6300947 B1	09-10-2001
		JP 2000066868 A	03-03-2000
WO 0039666 A	06-07-2000	DE 19962192 A1	06-07-2000
		FI 992746 A	28-06-2000
		GB 2347329 A	30-08-2000
		JP 2000194612 A	14-07-2000
		SE 9904687 A	29-06-2000
		WO 0039666 A1	06-07-2000
WO 0042531 A	20-07-2000	AU 2719800 A	01-08-2000
		WO 0042531 A2	20-07-2000

BNSDOCID: <WO___0221331A1_I_>